4L1-5in1

畳み込みニューラルネットワークを用いた表情表現の獲得

Feature Acquisition From Facial Expression Image Using Convolutional Neural Networks

西銘 大喜 *1 Taiki Nishime　遠藤 聡志 *2 Satoshi Endo　當間 愛晃 *2 Naruaki Toma　山田 孝治 *2 Koji Yamada　赤嶺 有平 *2 Yuhei Akamine

*1琉球大学大学院理工学研究科情報工学専攻
Graduate School of Information Engineering, University of The Ryukyus

*2琉球大学工学部情報工学科
School of Information Engineering, University of The Ryukyus

In this study, we carried out the facial expression recognition from facial expression dataset using Convolutional Neural Networks (CNN). In addition, we analyzed intermediate outputs of CNN. we analyzed the features that are included in the input. As a result, we have obtained a emotion recognition score of about 57%; two emotions (Happiness, Surprise) recognition score was about 70%. We also confirmed that CNN have learned the feature about Happiness from specific area of facial images. This paper details these experiments and investigations regarding the influence of CNN learning from facial expression.

1. はじめに

画像認識において, 畳み込みニューラルネットワーク (Convolutional Neural Networks : CNN) を用いた手法が, 高い性能を示すことが報告されている [1][2]. CNN は, 画像認識の一課題とされており, 表情認識にも応用されている.

表情には普遍性が存在し [3], 共通認識が可能だと考えられる. ただ, 表情は多くの種類が存在することから, 全てが等しく普遍性を持つのかは疑問である. Ekman らによれば, 幸福感, 驚き, 恐れ, 悲しみ, 怒り, 嫌悪 は普遍性を持つと定義されており [4], 表情認識においては, この 6 表情を扱うことが一般的である. 表情の種類は, 表情認識をアプリケーションへ応用する場合を考えると限定しやすいと考えている. 例えば, ユーザが笑顔になった Web コンテンツを記録するアプリケーションを考える場合は, 笑顔の認識が重要であり, 嫌悪や恐怖などの表情を考慮する必要がないのでは, と考えることができる.

CNN による高い画像認識精度は, 各層の処理で入力画像を低次元化する中で, 問題に適切な特徴量を抽出していることが, 強く影響していると考えられる. 学習される特徴量の検討に関連した研究として, 中間層出力から入力を再現する研究が挙げられる [5]. この研究は, 学習した特徴量の検討として有効であるが, 表情認識においては, 各表情の違いが複雑で小さいため, 入力を再現した際に表情間での区別が出来るほどの明瞭な違いを得ることは難しいと考えられる.

本研究では, 怒り, 嫌悪, 恐怖, 喜び, 悲しみ, 驚きの 6 感情に無表情を加えた 7 表情を対象に表情表現の学習を行い, CNN の入力に注目した分析を行う. 分析は, 入力画像の特定領域の値を変化させた場合の CNN の出力値を対象とし, 入力画像に含まれる表情の特徴量についての検討を行う.

2. 関連研究

これまで表情認識研究では, Facial Action Coding System(FACS) を用いる手法 [6][7], CNN を用いる手法の 2 種類がある. FACS は, 顔を Action Units と呼ばれる動作単位に分け, 各表情を定義し表情認識に利用される. FACS による各表情のラベル付けには, 専門的な知識が必要であり, 誰もが簡単にラベル付けが可能という訳ではない. このことに加え, CNN による高い認識精度を達成したことも影響して, CNN を用いて表情認識を行う VICTOR らの研究がある [8]. しかし, この研究では, 学習を終えた CNN に関する考察や, CNN が学習した特徴量への検討など, さまざまな疑問が残されている.

連絡先: 西銘 大喜, taiki_one@eva.ie.u-ryukyu.ac.jp

3. 畳み込みニューラルネットワーク

本研究では, 畳み込みニューラルネットワーク (CNN) で 9 層のネットワークを構築する. CNN は, 畳み込み層, プーリング層, 全結合層の 3 種類から構成されている. 各層の処理により, 入力画像を圧縮し特徴を抽出していく.

3.1 畳み込み層

畳み込み層では, 入力データの各画素値 (x_{ij}) と重みフィルタ (h_{pq}) の積和計算を行う. 重みフィルタのサイズ, 枚数は任意の数で指定可能である. 入力画像のサイズを $W \times W$ とし, 入力画素インデックスを $(i,j)(i=0,...,W-1,\ j=0,...,W-1)$ 重みフィルタのサイズを $H \times H$, 重みフィルタの画素インデックスは $(p,q)(p=0,...,H-1,\ q=0,...,H-1)$ とすると, 畳込み計算は式 (1) で表される.

$$u_{ij} = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x_{i+p,j+p} h_{pq} \tag{1}$$

また今回は, 重みフィルタを縦横方向に 1 画素ずつ動かしながら積和計算を行う.

3.2 プーリング層

プーリング層では, 図 1 のように入力データの中から特定領域を選択し, この中に含まれる画素値を選択する. 選択には幾つかの方法があり, 最大プーリングや平均プーリングなどがある. 名前の通り, 最大プーリングでは, 特定領域の中から最大値を選択し, 平均プーリングでは, 特定領域の平均値を求める.

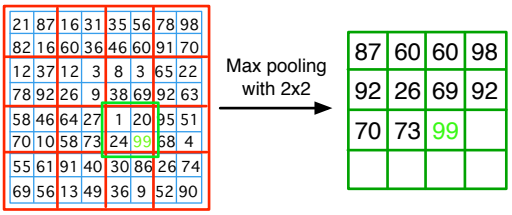

図 1: 最大プーリング, 入力データサイズを 8×8, プーリングサイズを 2×2 とした場合の例, 太線で囲まれた (3,3) の領域では, 最大値 99 が選択されている.

3.3 全結合層

全結合層は, 図 2 のような順伝播型ネットワークで表される. 次層の各ユニットの値 (Y_k) は, 前層のユニット全ての値 (X_i) と次層のユニットと前層を繋ぐ重みの値 (W_{ik}) の積和計算で求めた値を活性化関数 (s) を適用した値を利用する.

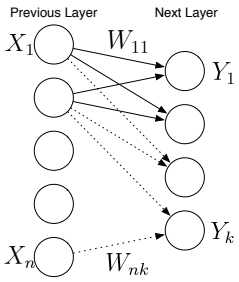

図 2: 全結合層, 各円は各層のユニット, 矢印はユニット通しを繋ぐ重みを表す. X_n と Y_k を繋ぐ重みを W_{nk} と表す. 各層の積和計算は式 (2) で計算される.

$$Y_k = s\left(\sum_{i=1}^{n} W_{ik} X_i\right) \tag{2}$$

4. 実験

本章では, 表情画像データセットと CNN を用いて表情表現の学習, 評価実験を行う.

4.1 実験環境

実験で使用した CNN, データセットの詳細を, それぞれ図 3, 図 4, 表 1 に示す. データセットは, Facial Expression Recognition 2013(FER-2013) データセット [9] を利用する. このデータセットは学習用, テスト用の 2 つに分かれており, 各画像はグレイスケール化されて, 48×48 のサイズで顔部分がトリミングされている. 今回は前処理として, Global Contrast Normalization(GCN) を行った. GCN では, 各画像の平均値と分散値を求め, 平均値で減算し, 標準偏差で除算し, 画素値の平均が 0, 分散が 1 になるように正規化を行う.

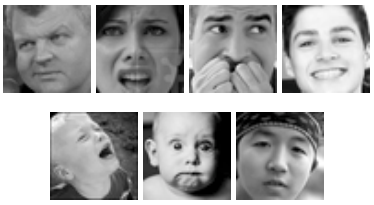

図 4: FER-2013 データセットの例, 左上から右下の順に, 怒り, 嫌悪, 恐怖, 喜び, 悲しみ, 驚き, 無表情となっている.

	怒り	嫌悪	恐怖	喜び	悲しみ	驚き	無表情	合計
Training	3993	436	4097	7212	4828	3171	4692	28698
Test	466	56	496	895	653	415	607	3588

表 1: FER-2013 データセットの詳細 (枚)

4.2 結果

実験結果を表 2 の混同行列に示す. 今回の評価実験では, 平均 56.9%の認識精度が得られた. 表 2 より, 笑顔と驚きの表情では 70%を超える精度が得られているが, 一方で怒りを悲しみに, 無表情を悲しみと誤認識している例も約 20%の誤認識率が確認できる. 嫌悪については学習用のデータ数が少ないことが大きく影響していると考えられる.

	怒り	嫌悪	恐怖	喜び	悲しみ	驚き	無表情
怒り	49.1	0.0	10.3	6.6	19.9	2.7	11.1
嫌悪	33.9	0.0	19.6	7.14	30.3	1.7	7.1
恐怖	11.8	0.0	35.8	4.8	26.4	7.6	13.3
喜び	5.4	0.0	3.3	76.6	5.2	2.3	6.9
悲しみ	10.8	0.0	12.8	7.3	51.1	0.9	16.8
驚き	5.5	0.0	7.2	5.5	4.0	71.5	6.0
無表情	9.3	0.0	5.6	9.2	21.5	1.3	52.8

表 2: 実験結果の混同行列 (単位:%)

また, 各データセットの中には, 正面以外の角度から撮影されたであろう表情画像や頭を傾けている表情画像も存在したが, それらのデータは特記する程に, 精度が悪い, 良い, というような傾向は確認できなかった.

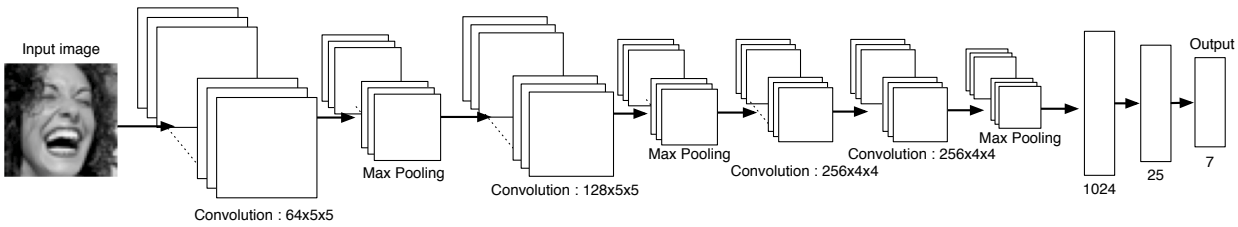

図 3: 実験に使用した CNN: 畳み込み層, プーリング層の後に全結合層を 2 層配置している. 数字は各層の次元数を表す.

5. 学習を終えた CNN の分析

5.1 分析手法

CNN により入力画像から得られる特徴量について, 各表情の帰属度を表す出力値の変化に注目し, 評価実験の結果も踏まえて検討する. ただし, 嫌悪の表情は分析対象から除き, CNN により正しく判断された画像に限定して分析を行った. 分析手法を以下に示す.

1. 入力画像を n 個の分析領域に分ける
2. 分析領域の中から 1 箇所選択し, 0.0, 0.2, 0.4, 0.6, 0.8, 1.0 の値でそれぞれ初期化し, マスク処理を行う.
3. 上記の処理を行った画像を入力し, CNN の笑顔の帰属度を示すユニットの出力値の変化を調べる

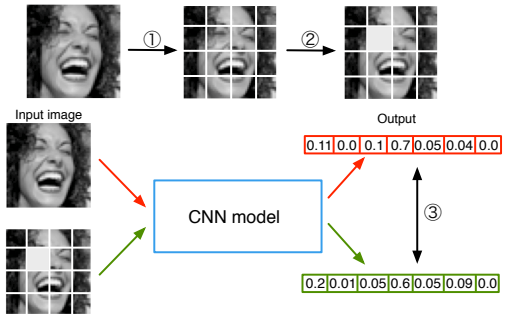

図 5: 分析手法 (n=16 の場合), それぞれ分析領域の初期化前と初期化後での, 出力値の変化に注目する.

5.2 分析結果と考察

認識精度が最も高い結果となった笑顔の表情について考察する. 笑顔表情の分析対象の画像例を図 6, 分析結果を図 7 に示す. また説明の都合上, 画像の分析領域に対し section1 から section16 と番号を振り分けた.

図 6: 分析画像例, 分析対象に多く見られた例

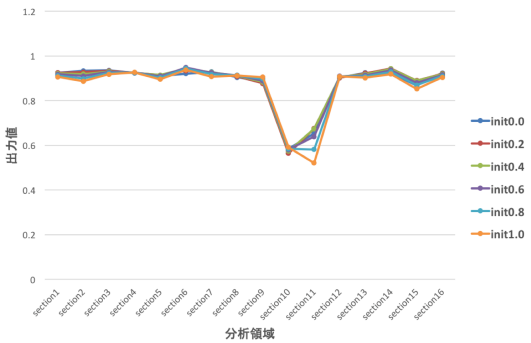

図 7: 分析領域と出力値平均 (笑顔)

図 7 は, 笑顔の帰属度を示す出力値の推移を表している. また分析対象の画像は x 軸は各分析領域を領域を表し, y 軸は笑顔の帰属度を示すユニットの出力値を表している. このグラフより, 分析対象が section10, 11 の場合に笑顔の出力値が低くなっていることが伺える. これは, 図 6 のように, 分析した画像の多くは section10, 11 の位置に口と鼻が位置しており, 笑顔の認識には, 目元, 口元の情報が大きく影響していると考えられる.

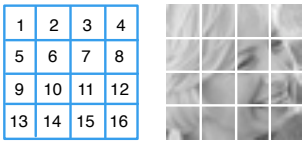

図 8: 分析画像例, 正面を向いていない例

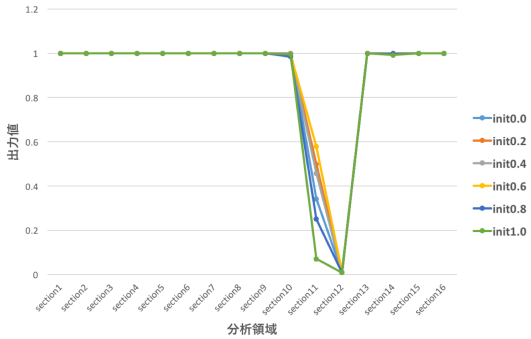

図 9: 分析領域と出力値 (笑顔)

次に, 分析対象画像の中であまり見られなかった, 顔が正面を向いていない笑顔画像での分析結果について述べる. 図 8 を対象とした場合の分析結果を示す (図 9). 先の結果 (図 7) とは異なり, section11, 12 で出力値が 0 付近を示していることがわかる. これは顔が正面を向いておらず, 口の位置が図 6 と図 8 で異なることが影響していると考えられる. この結果から, 画像内での口の位置に依らず, 口元の特徴が笑顔の認識に影響を与えていることが確認できた. このことからも笑顔の表情において, 口元の特徴を持つことを学習していることが確認できる.

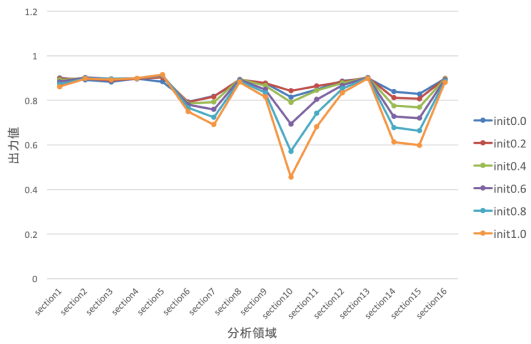

図 10: 分析領域と出力値平均 (驚き)

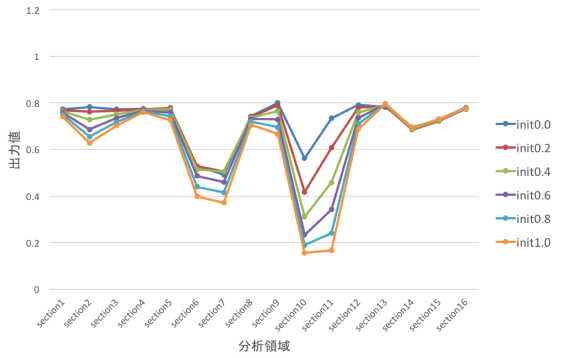

図 11: 分析領域と出力値平均 (怒り)

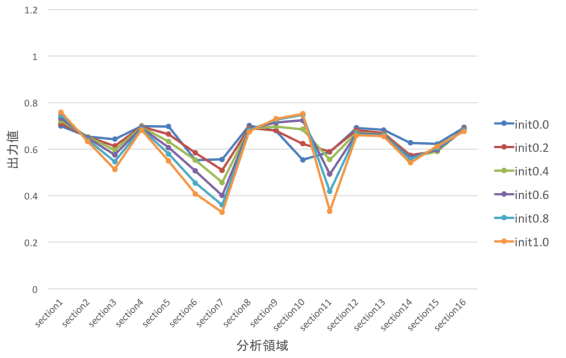

図 12: 分析領域と出力値平均 (悲しみ)

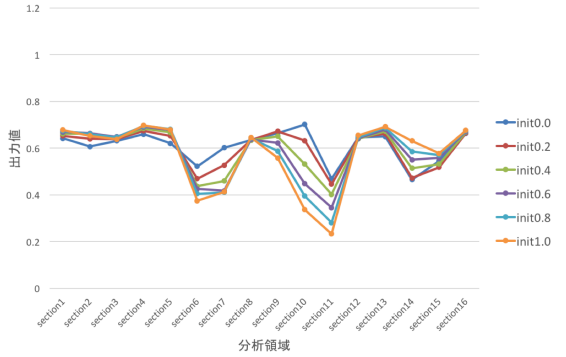

図 13: 分析領域と出力値平均 (恐怖)

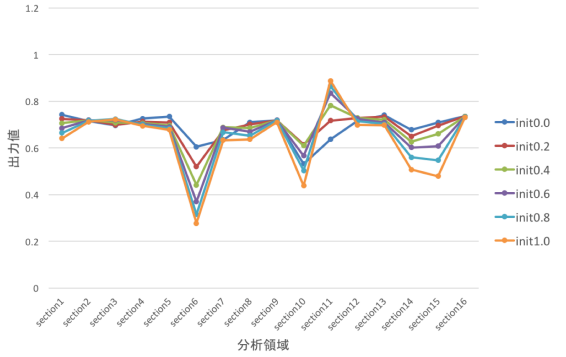

図 14: 分析領域と出力値平均 (無表情)

笑顔との比較として, 表情によって認識精度及び正例件数がそれぞれ異なる, 驚き, 怒り, 恐怖, 悲しみ, 無表情に関する考察を行う. 図 10, 11, 12, 13, 14 は, それぞれの分析結果を表している. それぞれの結果から, section6, 7, 10, 11 に影響を受け, 出力値平均が変化していることが確認できる. 笑顔 (図 7) や驚き (図 10) の表情とは異なり, 怒り, 悲しみ, 恐怖, 無表情の 4 種類は, section6, 7 にも影響を受け出力値が変化していることから, 口元の情報だけで認識することが難しく, 目元の情報も大きく影響しており, また, 表 2 の各表情の精度を踏まえて考えると, 表情毎の section6, 7, 10, 11 の違いを学習できていないとも考えられる.

6. まとめ

本研究では, CNN と表情画像データセットを用いて表情表現の学習を行い, 約 57%の精度が得られた. 学習後 CNN の分析から, CNN が口元の特徴を学習して, 表情表現の学習において有効に機能していることを確認した. 認識精度が低い表情は, さらなる分割数で処理を行った場合に, 各表情で異なる影響を受ける画像領域があれば, 該当箇所に強調するような前処理などを行う必要があり, 更なる検討が必要である.

参考文献

[1] ”Convolutional Neural Networks (LeNet) - DeepLearn- ing 0.1 documentation ”, LISA Lab. (2013)

[2] 岡谷 貴之, ”画像認識のための深層学習”, 人工知能, Vol.28, No.6, pp.962-974 (2013)

[3] 高木 幸子, ”コミュニケーションにおける表情および身体動作の役割”, (2005)

[4] Paul Ekman, W.V.Friesen, 工藤 力 (訳), ”表情分析入門”, (1987)

[5] K. Simonyan, A. Vedaldi and A. Zisserman, ”Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps”, Proc. ICLR Workshop (2014)

[6] 野宮 浩揮, 宝珍 輝尚, ”顔特徴量の有用性推定に基づく特徴抽出による表情認識”, (2011)

[7] Mengyi Liu et al. ”AU-aware Deep Networks for Facial Expression Recognition”, (2013)

[8] VICTOR-EMIL NEAGOE, ANDREI-PETRU B UTF0102RAR, NICU SEBE, PAUL ROBITU,”A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions”, Recent Advances in Image, Audio and Signal Processing, pp.93-98 (2013)

[9] Goodfellow, Ian J et al.“ Challenges in Representation Learning: A report on three machine learning contest ”Neural Information Processing. Springer Berlin Heidel- berg, (2013)